Cuadernos de Investigación Social
Serie Textos Fundamentales

Colección dirigida por Manuel Mora y Araujo
(Departamento de Sociología de la Fundación Bariloche)

Francis Korn
Paul Lazarsfeld
Allen H. Barton
Herbert Menzel

Conceptos y variables en la investigación social

Selección de Francis Korn y Manuel Mora y Araujo

Ediciones Nueva Visión
Buenos Aires

Queda hecho el depósito que previene la ley 11.723
Impreso en la Argentina / Printed in Argentina

Viamonte 494, Buenos Aires, República Argentina.

1

Francis Korn

Introducción

El significado del término “variable” en sociología

0. Los tres artículos que componen este cuaderno constituyen, a nuestro juicio, una buena introducción al problema de la investigación empírica en sociología. El tema de la operacionalización de conceptos al que se refieren constituye el primer paso hacia la comprensión del diseño experimental y de las técnicas de análisis de datos.

Dado que estos artículos han sido seleccionados de distintas fuentes y que, por tanto, no forman una secuencia ordenada y completa, dedicaremos esta introducción a una revisión del problema. A tal efecto, trataremos de presentar una elucidación del término "variable" y una clasificación de sus distintas dimensiones.

1. *La noción de "variable".* El término "variable" está revestido de una suerte de prestigio derivado de su origen matemático y se lo emplea habitualmente para connotar científicamente cualquier discurso, o bien se lo teme como algo cuyo significado es profundo e inalcanzable. En realidad, su uso indiscriminado contribuyó a que perdiera su preciso significado original y se convirtiera en un sinónimo de "aspecto" o "dimensión". Lo que normalmente se llama "variable" en el discurso sociológico no significa otra cosa que un aspecto o dimensión de un objeto.

Una "dimensión de variación", una "*variante*" o una "variable", designa en la jerga sociológica un aspecto discernible de un objeto de estudio. En cualquier proposición sociológica, estos términos designan las partes no relacionales de la proposición. Así, en una proposición del tipo: "Las frecuencias de ciertos cuadros neuróticos están relacionadas con el status socio-económico", "cuadros neuróticos" y "status socio-económico" constituyen las variables de la proposición porque proponen los aspectos individuales que tendrán que ser medidos para poner a prueba la relación enunciada.

Pero cualquier persona que intente someter tal proposición a una prueba empírica para comprobar su validez va a tropezar con una serie de dudas. Por ejemplo: qué rasgos específicos de la personalidad considerará trastor-

nos neuróticos, qué dimensiones del status social de los individuos serán relevantes para ubicarlos en status socio-económicos determinados, cuáles serán los criterios según los cuales diferenciará distintos grupos de status y, finalmente, qué agrupamiento de datos será el que le indique que efectivamente la relación entre ambos fenómenos existe.

En los tres primeros casos, la solución provendrá del marco teórico particular que el investigador elija para definir estos términos y de sus posibilidades prácticas de operacionalización y medición; en el último, requerirá primero la determinación del tipo de relación (determinista o probabilística propuesta y de las medidas y técnicas elegidas para su prueba.

Pero sea cual fuere la definición teórica adoptada, en el caso de los trastornos neuróticos y el status socio-económico de nuestro ejemplo, éstas deberán ser tales que permitan:

(i) definir ambos conceptos de modo que sea posible encontrar grupos de individuos que se diferencien claramente respecto a los tipos de neurosis y a su status socio-económico, y

(ii) comparar los grupos de individuos diferenciados según una de las dos variables en cuanto a las frecuencias de los distintos grupos de la otra.

En otros términos, el significado completo de la palabra “variable”, tal como es usada en ciencias sociales, contiene no sólo la connotación de “aspecto” o “dimensión”... de un fenómeno, sino también la propiedad de estos aspectos o dimensiones de asumir distintos valores.

2. *Operacionalización de las variables.* El tipo de definición (real o nominal) y la particular connotación de los términos incluidos en una proposición sociológica dependen del completo cuerpo teórico en que esta proposición está contenido. Pero si una proposición tiene el status de hipótesis científica, es decir, si su verdad

depende de su posibilidad de refutación empírica, sus variables deben ser traducidas a conceptos mensurables. Normalmente, las proposiciones sociológicas se expresan en términos que se refieren a cualidades de objetos o relaciones entre ellos, y no a perceptos. Corresponden a la categoría de “hechos sociales no manifiestos” en la terminología de Durkheim. Como la prueba empírica implica mediación, es necesario estudiar estos hechos sociales no manifiestos por los hechos sociales manifiestos que los representan. Es decir, es necesario definir las variables teóricas contenidas en una hipótesis en términos de variables empíricas. A estas últimas se las llama *indicadores.* La elección de los indicadores de una variable depende también, obviamente, del marco teórico con el que se está trabajando y de una serie de decisiones referidas a las posibilidades técnicas de medición. La cantidad de indicadores requeridos para representar la totalidad del significado de una variable depende a su vez de la cantidad de aspectos discernibles de esa variable, aspectos a los que se da el nombre de “dimensiones” para denotar su propiedad de ser parte de una totalidad mayor y que generalmente se enumeran en la definición real de la variable. Así el proceso lógico en la operacionalización de una variable requiere los siguientes pasos:

(i) definición nominal de la variable a medir;

(ii) definición real: enumeración de sus dimensiones;

(iii) definición operacional: selección de indicadores.

En el caso del “status socio-económico”, si se considera que sus aspectos relevantes para la medición son el económico, el de prestigio y el educacional, estas tres dimensiones podrían ser medidas por otros tantos indicadores como: ingresos individuales mensuales, ocupación (previo conocimiento del prestigio relativo de las ocupaciones en la particular sociedad en la que se está operando) y grado de educación formal alcanzado. Obviamente, la cantidad de operaciones que hay que realizar con cada una de las dimensiones de una variable para volverla mensurable depende de su distancia al plano

empírico. También hay que tener en cuenta que la cantidad de Indicadores que se use para representar una variable dependerá no sólo de su complejidad conceptual, sino de la cantidad de pruebas empíricas que requerirá su validación. A fin de ejemplificar este último punto, es claro que para medir la dimensión educacional de la variable "status socio-económico" puede bastar el conocimiento del grado de educación formal de un individuo; en cambio la dimensión de prestigio de la misma variable requerirá por lo menos los siguientes indicadores: nombre de la ocupación, posición dentro de la ocupación, tareas que cumple el individuo y, como dijimos más arriba, el conocimiento de la escala de prestigio de las ocupaciones en el lugar donde se realiza la medición.

La medida compleja que se obtiene combinando los valores obtenidos por un individuo en cada uno de los indicadores propuestos para la medición de una variable se denomina índice. La diferencia entre un índice y un indicador es entonces de grado. Un índice es un complejo de indicadores de dimensiones de una variable y constituye, por lo tanto, el indicador total de una variable compleja.

Lo dicho hasta aquí puede servir de introducción a dos de los artículos contenidos en este libro. El capítulo 1, "Nacimiento y desarrollo de las variables", de Lazarsfeld, trata del proceso por el cual el investigador selecciona indicadores para sus variables, partiendo de una primera imagen mental de la dimensión que quiere medir, para llegar a la construcción del índice. Discutiendo este proceso, Lazarsfeld examina algunas decisiones críticas que deben enfrentarse: la especificación del concepto en sus principales dimensiones o aspectos componentes —que pueden tener un origen tanto empírico como lógico—, la selección de los indicadores a partir de un conjunto potencialmente infinito de referentes empíricos, y la decisión de combinarlos de una determinada manera para formar un índice final. Una de las propiedades de los Indicadores que Lazarsfeld examina especialmente es la de su intercambiabilidad. Este punto es particularmente importante en el proceso de validación de definiciones

operacionales, porque la propiedad de la intercambiabilidad constituye uno de los fundamentos de la construcción de índices compuestos por varios indicadores.

El proceso de construcción de variables compuestas es tratado más en profundidad en el capítulo 2, de Barton, quien discute los diversos aspectos distintivos de un espacio de propiedades, o atributos, o variables sociológicas, proporcionando así las herramientas esenciales para la construcción de variables. Si bien las operaciones que Barton discute en su trabajo no siempre son explícitamente realizadas por el investigador, los principios generales subyacen a todo proceso de construcción de variables. La idea de espacio de atributos corresponde a una representación geométrica de la dimensión representada por un indicador, de tal manera que varios indicadores simultáneos definen un espacio complejo en el cual cada unidad de análisis puede ser representada por un punto. Esta conceptualización es especialmente útil para aclarar algunas operaciones que subyacen al proceso de construcción de índices en la investigación empírica, a saber: la simplificación de un espacio o emergencia de un nuevo aspecto de una dimensión, o lo que Barton llama "reducción"; y la operación por la cual el espacio es desarrollado a partir de sus componentes iniciales, o "substrucción".

3. *Unidades y propiedades que puede designar una variable.* El capítulo 3, "Sobre las relaciones entre propiedades individuales y colectivas", de Lazarsfeld y Menzel, aporta un importante intento de clasificación de variables sociológicas. Los criterios de clasificación en este caso son dos:

a) referido a las *unidades* en que han de ser medidas las propiedades que definen una variable y que pueden ser:

   (i) *miembros* (unidades que forman parte de otras unidades más amplias), y

(ii) *colectivos* (unidades que pueden ser descompuestas en otras menores).

b) referido a *propiedades* de esas unidades o a *relaciones* entre ellas y que pueden ser:

(i) propiedades que dependen de las unidades mismas;

(ii) propiedades que dependen de las unidades mayores que integran una unidad (para los miembros) o de las unidades menores en que puede descomponerse una unidad (para los colectivos);

(iii) relaciones entre los miembros;

(iiii) relaciones entre los miembros y el colectivo de que forman parte.

Así, cruzando los dos criterios, resultan las siguientes propiedades:

		unidades	
		miembros	colectivos
propiedades que dependen de los	miembros	absolutas	analíticas
	colectivos	contextuales	globales
relaciones entre	miembros	relacionales	estructurales
	miembros y colectivos	comparativas	

Esta clasificación parece en principio derivada de una pura vocación taxonómica de los metodólogos, pero una vez ordenada presta real utilidad en la elaboración e interpretación de los datos. La distinción entre propiedades analíticas de los colectivos y absolutas de los miembros provee

de una conceptualización necesaria para el análisis de la falacia ecológica,[1] donde precisamente son confundidas. Respecto a esta clasificación, sin embargo, quedarían por discutir algunos puntos oscuros. Para no extender demasiado esta introducción y considerar algunas otras clasificaciones, discutiremos aquí sólo el problema de la definición de "miembros" y "colectivos". Lo que no queda claro en el trabajo de Lazarsfeld y Menzel es si la distinción entre ambos términos es puramente lógica o si se trata de una distinción basada en las definiciones sustantivas de las unidades. No queda claro si un colectivo es cualquier conjunto lógico (como por ejemplo un grupo de personas agrupadas por la sola razón de que todas tienen la misma edad) o es sólo un conjunto tal que posea la propiedad de poder ser tratado como un "actor social" (como sería el caso del conjunto de personas que componen un grupo institucionalizado: una familia, un partido político). La ambigüedad de la distinción en el texto podría salvarse si se definen como "colectivos" todos aquellos conjuntos que posean por lo menos una propiedad "global", es decir, que puedan ser definidos por alguna propiedad ajena a las propiedades de sus miembros. Siempre que se elimine de entre ellas la de "tamaño" o "número de miembros", se eliminan de la definición de "colectivo" los simples agregados estadísticos como los grupos de edades, y sólo caerán dentro de la definición aquellos conjuntos a los que se puede considerar como sociológicamente existentes.

4. *Medición y clasificación.* Habitualmente las variables son clasificadas de acuerdo con el nivel en que permiten medir los objetos. El continuo nivel de medición va desde las mediciones muy débiles hasta las mediciones fuertes en términos del grado de cuantificación al-

[1] Una excelente y completa discusión del problema de las falacias ecológicas, o falacias "de nivel equivocado" se encuentra en Johann Galtung, *Teoría y método de la investigación social,* Buenos Aires, Eudeba, 1966.

canzable. Aquí no entraremos a analizar la clasificación de las variables en términos del nivel de medición, sino que abordaremos algunos aspectos lógicos de una variable vista como un *sistema clasificatorio*. Para ello consideraremos primero el significado del término "categoría" en el proceso de medición.

Más arriba, cuando tratábamos de definir qué es una "variable", dijimos que una inseparable connotación de la misma es su capacidad de asumir distintos valores. El conjunto de valores en que se divide una variable es denominado indistintamente "sistema de valores" o "sistema de categorías". Estos sistemas no son fijos para una variable dada, ya que dependen tanto de la naturaleza de la variable como de las decisiones del investigador que las manipula de acuerdo con la importancia de la variable en la prueba que está realizando, el particular contexto en que trabaja y el tipo de técnicas y marco teórico que utiliza. Para volver al ejemplo mencionado, si se está trabajando con una hipótesis que contiene una variable denominada "trastornos neuróticos", la primera decisión con respecto al sistema de categorías que se va a emplear para medirlos dependerá de la distribución de dichos trastornos en la población donde se realizará la medición. (Obviamente, ésta será la primera decisión a partir de una clara definición conceptual del término.) La más gruesa categorización al respecto será la que divida el grupo de individuos en que se realizará la medición en "neuróticos" y "no neuróticos". Es decir, la categorización más básica de una variable es la que distingue entre la presencia y la ausencia de la propiedad que ella enuncia. Las variables que sólo pueden subdividirse de este modo, o clasificando distintos tipos cualitativamente diferentes de una propiedad (en nuestro caso un sistema que las dividiera en "neurosis obsesiva, histérica, fóbica y paranoica", por ejemplo), pero no en sistemas donde cada categoría represente un grado diferente de la misma propiedad, se denominan generalmente "atributos" o "variables cualitativas" o "variables nominales".

Según Torgerson,[2] estas variables no constituyen la base de ninguna medición, ya que sólo clasifican distintas propiedades y no distinguen entre distintos grados de una misma propiedad (en nuestro caso, la única posibilidad de este tipo sería distinguir entre diversos grados de gravedad de la neurosis en general o de cada una de sus clases). Pero nosotros consideramos, siguiendo por ejemplo a Stevens,[3] que este tipo de variables constituye el más bajo nivel de medición de una propiedad. Esta decisión se basa en el hecho de que cuando se categoriza según grados de una propiedad también se está clasificando y de que, en sociología, la frecuencia de este tipo de variables es muy grande.

Pero sea nominal o cuantitativo, el sistema en que se subdivide una variable será lógicamente válido si considera que *la relación que existe entre un conjunto de individuos y un conjunto de valores de una variable es la siguiente:*

> *Si A es el conjunto de individuos a clasificar* (por ejemplo el conjunto de individuos de una sociedad)
>
> *y B el conjunto de valores de una variable* (por ejemplo el conjunto de categorías de la variable "clase social")
>
> la clasificación de los elementos del conjunto A según los elementos del conjunto B será *una función del primer conjunto en el segundo* (función de A en B), *tal que a cada elemento de A le corresponda uno y sólo uno de B.*
>
> (En nuestro caso, cada individuo será clasificado en una y sólo una clase social.)

De modo que la clasificación de cualquier conjunto se realiza de la siguiente manera:

[2] *Theory and Methods of Scaling,* Nueva York, Wiley, 1962.
[3] *Handbook of Experimental Psychology,* Nueva York, Wiley, 1962, pp. 1-30.

Dos elementos del conjunto A están en el mismo subconjunto de A (son iguales en lo que respecta a A) sí y sólo si van a parar (por la función) al minino subconjunto de B.

(Dos individuos pertenecerán al mismo sector de la sociedad si se los clasifica en la misma clase social.)

Lo antedicho resume dos de las propiedades de cualquier sistema de medición: (i) que sus categorías sean mutuamente excluyentes y (ii) que el sistema permita clasificar a todos los individuos de la población considerada.

Restaría comentar una tercera propiedad de un buen sistema de categorías: su relevancia para el contexto en que va a ser aplicado. Daremos algunos ejemplos que pueden mostrar con mayor claridad la importancia de esta dimensión. Toda variable tiene una "amplitud de variación" teóricamente mucho mayor que la que resulta relevante en cada circunstancia específica. Un buen sistema de categorías, además de ser lógicamente correcto, debe servir a un propósito determinado en la investigación y ser económico. Algunos ejemplos pueden hacer más clara la importancia de este punto. Tomemos el caso de los conflictos socio-políticos, que pueden ser clasificados, de acuerdo con sus motivos o fines manifiestos, en económicos, políticos y étnico-religiosos. Esta clasificación puede ser lógicamente correcta, pero sería poco relevante en un país como la Argentina, donde prácticamente no existen conflictos étnico-religiosos, y donde la mayoría de los conflictos son económicos o políticos. Sería de mayor interés refinar más la clasificación partiendo de las categorías más frecuentes. Para tomar otro caso, se puede clasificar a inmigrantes por su nacionalidad; una clasificación basada en el continente de origen ("latinoamericanos, europeos, africanos y asiáticos") es poco relevante para el contexto argentino, obviamente. La mayoría de los inmigrantes caerán en las dos primeras categorías, dejando las dos últimas prácticamente

vacías. Tal sistema sería antieconómico y poco adecuado, ya que permitiría establecer muy poca diferencia entre los inmigrantes. Sacrificar la parte de variación de la variable que no es relevante y profundizar más la parte relevante es un principio importante que complementa la corrección lógica del sistema clasificatorio.

5. Con estas consideraciones queda introducida la noción de "variable" tal como se la utiliza en la investigación social. Las técnicas de construcción de variables, como la formación de índices y escalas, y las técnicas de validación de indicadores –por las cuales se pueden hacer inferencias legítimas del nivel manifiesto al latente– no se examinan aquí. Estas técnicas son diversas y se desarrollan continuamente, pero los principios generales subyacentes a ellas se modifican en mucho menor medida. Los tres artículos que siguen fueron seleccionados precisamente con la intención de ofrecer exposiciones claras sobre algunos de tales principios, y en la creencia de que su estudio y discusión en profundidad pueden resultar de igual o mayor importancia para la investigación empírica que el dominio de las diversas técnicas más particulares.